# Information Compression by Multiple Alignment, Unification and Search as a Framework for 'Intelligent' Computing*

Gerry Wolff,
*School of Informatics, University of Wales Bangor,*
*Dean Street, Bangor, LL57 1UT, UK.*
*E-mail: gerry@sees.bangor.ac.uk.*
*Fax: +44 1248 361429.*

## Abstract

This paper presents an overview of the idea that several aspects of 'intelligent' computing may be understood as *information compression by multiple alignment, unification and search* (ICMAUS). The meanings of these terms are explained.

Aspects of this framework are realised in the SP61 computer model, which is described in outline.

Brief descriptions are given of how the framework may accommodate such things as natural language processing, 'fuzzy' pattern recognition, best-match information retrieval, classes, subclasses and inheritance of attributes, probabilistic 'deductive' and abductive reasoning, chains of reasoning, default values and nonmonotonic reasoning, the phenomenon of 'explaining away', and unsupervised inductive learning. Examples are illustrated with output from the SP61 model.

*Key words:* information compression; multiple alignment; probabilistic reasoning

## 1 Introduction

For several years, I have been developing a conceptual framework intended to integrate and simplify a range of concepts in computing, including those associated with human-like 'intelligence' in computing such as machine learning [2, 15, 16], probabilistic reasoning [9, 2], 'fuzzy' pattern recognition and information retrieval [15], natural language processing [8], and others. This framework is partially realised in a computer model, SP61.

As a unifying thread, this work has been guided by the conjecture that *all kinds of computing and formal reasoning may be understood as information compression by pattern matching, unification and search* [19, 18, 17, 16, 15, 14, 13].

In recent work, this conjecture has been refined to the conjecture that *all kinds of computing and formal reasoning may be understood as* **information compression** *by* **multiple alignment**, **unification** *and* **search** (ICMAUS): [12, 11, 10, 9, 8].

By means of examples, this paper aims to present a brief, informal overview of this work, focussing mainly on aspects that relate to 'intelligent' analysis of data.

For the sake of clarity and to save space, the examples are quite small. But the framework and the SP61 model are capable of handling much larger examples. An analysis of the computational complexity of the model shows that its scaling properties, both for time complexity and for space complexity, are reasonably good [8].

## 2 The Overall Framework

In its most general form, the ICMAUS framework is envisaged as a system for the unsupervised inductive learning of grammar-like structures that works like this:

1. Starting with little or no knowledge of the 'world', the system receives raw data from the world via its 'senses'. These data are designated 'New'.

2. As each portion of New is received, the system tries to compress it as much as possible by matching it against stored patterns and encoding it in terms of patterns or structures that the

*Paper presented at the International ICSC Congress on Computational Intelligence: Methods & Applications (CIMA 2001), June 2001, Bangor, UK.

system already knows. If for example, the system knows the pattern "information compression" and has given it the code "IC", then whenever this pattern appears in New, it may be abbreviated as "IC". A 'good' match between patterns is one that yields a relatively large compression of New.

3. Portions of New that can be encoded in terms of stored knowledge, are stored in their encoded form. Every other portion is stored in raw form but is given a new code by the system for possible use in the encoding of New information in the future. All stored knowledge is designated 'Old'.

In broad terms, this incremental scheme is similar to the well-known and widely-used Lempel-Ziv algorithms for information compression.

What is different about the ICMAUS scheme is its exploitation of heuristic search to find relatively 'good' partial matches between patterns and the way a concept of 'multiple alignment' has been developed to support the encoding New information in a hierarchy of 'levels', as will be seen in examples below. Principles of Minimum Length Encoding (MLE, [6, 7, 4]) are an explicit foundation for this development.[1]

# 3 The SP61 Model

The SP61 computer model is described most fully in [8] and described in outline in [10]. At the heart of the model is a process for finding 'good' partial matches between two patterns which is essentially a form of dynamic programming (see, for example, [5]) but with some advantages over standard methods, including:

- The ability to process arbitrarily long patterns and
- The ability to find many alternative matches between patterns.

A relatively full description of the process may be found in [15] (see also [14]).

The 'dynamic programming' in SP61 is applied iteratively in such a way that the system can find alignments amongst arbitrarily large numbers of patterns, not merely two. As will be seen, this capability means that it can build structures that reflect arbitrarily deep hierarchies in cognitive structures.

So far, the SP61 model has been developed only for stage 2 of the scheme described in Section 2, as will be seen below. Work is in progress now to generalise the model to achieve the remaining aspects of the overall framework as described in Section 2.

[1]In brief, the key idea in MLE is that, in the unsupervised learning of a grammar, one should aim to minimise $(G + E)$, where $G$ is the size of the grammar in bits and $E$ is the size of the sample text (in bits) after it has been encoded in terms of the grammar.

S → NP VP
NP → D N
D → t h a t
D → o n e
N → t r a i n
N → b i c y c l e
VP → g o e s
VP → s t o p s

Figure 1: A fragment of the grammar of the syntax of English.

S NP #NP VP #VP #S
NP D #D N #N #NP
D t h a t #D
D o n e #D
N t r a i n #N
N b i c y c l e #N
VP g o e s #VP
VP s t o p s #VP

Figure 2: A fragment of the grammar of the syntax of English translated into ICMAUS *patterns*.

# 4 Natural Language Processing

A good introduction to the nature of the current system and its capabilities is the way it can be used for natural language processing. Consider a simple grammar like the one shown in Figure 1.

Given a grammar like this, a sentence like 't h a t b i c y c l e g o e s' may be parsed as 'S(NP(D(t h a t)N(b i c y c l e))VP(g o e s))', or an equivalent representation as a tree.

In the ICMAUS scheme, the grammar shown in Figure 1 may be translated into a set of *patterns*, like those shown in Figure 2.

Given the sentence 't h a t b i c y c l e g o e s' in New and the patterns in Figure 2 in Old, SP61 forms a variety of alignments, the best one of which is shown in Figure 3.[2] This alignment may be unified to form the pattern 'S NP D t h a t #D N b i c y c l e #N #NP VP g o e s #VP #S', which may be seen to be equivalent to the parsing shown earlier.

In [8], more sophisticated examples may be found showing that the ICMAUS framework is nor merely a novel version of context-free phrase-structure grammar (CF-PSG) but that it has the 'power' of context-sensitive systems and that it allows linguistic structures to be represented in a manner which is arguably simpler and more transparent than in existing systems.

In the same source, it may be seen that *the same system* may be used *without modification* to achieve the *production* of language as well as the parsing of language.

### 4.1 Evaluation of Alignments

Alignments like the one shown in Figure 3 are evaluated in terms of the compression of New that may be achieved by encoding it in terms of the patterns from Old that appear in the alignment.

Given the close relation between information compression and probability,[3] each alignment may also be evaluated in terms of probability. SP61 makes the necessary calculations using information which is provided about the frequency of occurrence in some domain of each pattern in Old.

## 5 'Fuzzy' Pattern Recognition, Scene Analysis and Best-Match Information Retrieval

The 'dynamic programming' feature of the ICMAUS framework (and the SP61 model) means that it is just as much at home with partial matches between patterns as with full ('exact') matches between patterns. This is the key to the system's ability to recognise patterns in a 'fuzzy' manner—like systems for spelling checking or, indeed, human capabilities for recognising patterns and objects.

Figure 4 shows, in order of their compression scores, the three best alignments found by SP61 with

```
0 i n p f   r m a t i x n 0
  | |   |   | | | | |   |
1 i n   f o r m a t i o n 1

0 i n p f r m a t   i x n 0
      |   |   |     |   |
1     p a r   a f f i   n 1

0 i n p         f r m a t i x n 0
      |         |         |   |
1     p a r a f f         i   n 1
```

Figure 4: The three best alignments found by SP61 with 'i n p f r m a t i x n' in New and a small dictionary in Old.

the mis-spelled word 'i n p f r m a t i x n' in New and a small dictionary of correctly spelled words in Old. The first alignment shows how the correctly-spelled word may be found despite additions, omissions and substitutions in the query word. Other examples may be found in [9].

### 5.1 Scene Analysis and Best-Match Information Retrieval

The system's ability to find good partial matches between patterns suggests that it may be generalised in the future to imitate the way we can recognise objects in a typical scene despite the fact that most objects are partially obscured by others.

As with pattern recognition, the system's ability to find good partial matches between patterns means that it can achieve best-match retrieval of information from a database—finding good partial matches between a 'query' pattern and zero or more patterns in a database. Examples and discussion may be found in [15].

## 6 Classes, Subclasses and Inheritance of Attributes

A prominent feature of human cognition and perception is the grouping of entities into hierarchies of classes with 'inheritance' of attributes from higher-level classes to lower.

Figure 5 shows how we might recognise a creature as a cat from knowing that it purrs and, at the same time, we may recognise it as a mammal and a vertebrate. Knowing that it is a cat means that we can infer that it has retractile claws, knowing that it is a

[2] By convention in this and all other alignments, New is shown in row 0 of the alignment and patterns from Old are shown below it.

[3] Probabilities have a key role in most methods for information compression. And calculated values for compression may be translated into probabilities (see, for example, [1]).

```
0        t h a t      b i c y c l e           g o e s        0
         | | | |      | | | | | | |           | | | |
1        | | | |    N b i c y c l e #N        | | | |        1
         | | | |    |               |         | | | |
2   NP D | | | | #D N               #N #NP    | | | |        2
    |  | | | | | |                      |     | | | |
3   |  | | | | | |                      |  VP g o e s #VP    3
    |  | | | | | |                      |  |           |
4 S NP | | | | | |                     #NP VP         #VP #S 4
       | | | | | |
5      D t h a t #D                                          5
```

Figure 3: The best alignment found by SP61 with 't h a t b i c y c l e g o e s' in New and the patterns shown in Figure 2 in Old.

mammal means that we can infer that it is furry and warm-blooded, and knowing that it is a vertebrate means we can infer that it has a backbone.

# 7 Probabilistic Reasoning

Figure 5 illustrates the way in which alignments can support reasoning. From the fact that an animal purrs we infer that it is a cat (with the attributes of cats) and also that it is a mammal and a vertebrate (with the attributes of those categories). In general, the symbols in Old and New are the propositions which are the basis of reasoning: In the best alignment or alignments that are found, each symbol in a pattern from Old that is not aligned with any symbol in New represents an inference made by the system.

In general, these inferences are probabilistic because there is a probability associated with each alignment. A relatively full discussion of the way probabilistic inferences can be modelled in the ICMAUS framework may be found in [9].

## 7.1 Probabilistic 'Deduction'

Figure 6 illustrates the way in which the ICMAUS framework may be used for probabilistic 'deduction': from knowing that Tweety is a bird we can infer that it can fly (and also that it has other attributes of birds such as feathers, beak etc). The probability associated with the alignment represents the probability of the corresponding inferences. In this example, the probability is 1.0 because the alignment shown is the only one that matches all the symbols in New.

In reality, of course, we know that some birds cannot fly. But this information was not recorded in the small database of patterns supplied to SP61 in this case. The way SP61 can handle a more true-to-life example is described briefly in Section 7.4, below.

## 7.2 Abduction

The nice thing about this framework is that it works just as well in a 'backwards' abductive style as in the 'forward' style just shown. Figure 7 shows how a knowledge that Tweety can fly leads the system to infer that he or she could be a bird or, alternatively, that he or she could be bat. Each of these possibilities has an associated probability, calculated by SP61 to be 0.8 in the first case and 0.2 in the second.

## 7.3 Chains of Reasoning

Apart from the kinds of one-step reasoning described in Sections 7.1 and 7.2, we can of course reason in 'chains': "If A then B, if B then C" and so on. The ICMAUS scheme lends itself very well to this kind of reasoning as can be seen schematically here:

```
A B
  |
  B C
    |
    C
```

Apart from simple chains like the one shown, the ICMAUS framework has the flexibility to accommodate more subtle kinds of composite reasoning, as described in [9].

```
0                                                       purrs                   0
                                                          |
1 cat mammal                                         #m purrs retractile-claws #ct 1
       |                                              |
2     mammal vertebrate          #v furry warm-blooded #m                       2
                |                 |
3            vertebrate backbone #v                                            3
```

Figure 5: A simple example showing how a class hierarchy with inheritance of attributes may be accommodated in the ICMAUS framework.

```
0 bird      Tweety                                                       0
   |          |
1  |   name Tweety #name                                                 1
   |    |            |
2 bird name        #name canfly wings feathers beak crop lays_eggs ... #bird 2
```

Figure 6: The best alignment found by SP61 with 'bird Tweety' in New and a small database of patterns about different kinds of animals in Old.

```
0           Tweety       canfly                                          0
              |            |
1      name Tweety #name   |                                             1
        |            |     |
2 bird name        #name canfly wings feathers beak crop lays_eggs ... #bird 2

0          Tweety          canfly                    0
             |               |
1     name Tweety #name      |                       1
       |            |        |
2 bat name        #name fur canfly eats_insects ... #bat 2
```

Figure 7: The two best alignments found by SP61 with 'Tweety canfly' in New and a small database of patterns about different kinds of animals in Old.

### 7.4 Default Values and Nonmonotonic Reasoning

As noted above, it is not true to life to assert that all birds can fly. To be more realistic, the knowledge that Tweety is a bird should lead to an inference that, *probably*, Tweety can fly. If, subsequently, we learn that Tweety is a penguin, it should be possible to reverse our default assumption that Tweety can fly and reach the conclusion that Tweety cannot fly.

The way in which this kind of 'nonmonotonic' reasoning with default values can be modelled in the ICMAUS framework is described in [9], illustrated with an example from SP61.

### 7.5 Explaining Away 'Explaining Away'

In the words of Judea Pearl [3, p. 7], the phenomenon of 'explaining away' may be characterised as: "If A implies B, C implies B, and B is true, then finding that C is true makes A *less* credible. In other words, finding a second explanation for an item of data makes the first explanation less credible." (his italics). As an example, one might receive a telephone call at work to say that one's burglar alarm (at home) has sounded. Normally, this would mean a burglary at the house. But, if the burglar alarm is sensitive to earthquakes and if there had been a radio announcement that an earthquake had occurred recently, one would probably conclude that this was the cause of the burglar alarm going off.

Pearl describes how this kind of 'explaining away' can be modelled using causal networks. The ICMAUS framework can also model this kind of reasoning, as described in [9] with an example from the SP61 model.

## 8 Machine Learning

The ICMAUS framework outlined in Section 2 has its origins in research developing computer models of the unsupervised learning of language by children [21, 20]. With respect to syntax at least, it seems that such learning may, to a large extent, be understood as information compression by the unification of matching patterns. These ideas rest on the twin foundation of MLE principles [6, 7, 4] and distributional principles proposed in the structuralist tradition of linguistics.

Although working models of unsupervised learning were developed in the earlier research on language learning, they did not meet a number of requirements for a generalised model of different kinds of computing. Hence, most of the effort to date in this latter development has been devoted to aspects of computing other than learning and it is only recently that the work has come full circle to develop the SP61 model for unsupervised learning as outlined in Section 2.

Although it is not yet possible to demonstrate unsupervised learning in an ICMAUS computer model, it is possible to see in general terms how distributional principles combine with MLE principles to achieve learning.

Consider, for example, how a child might begin to construct a grammar for his or her native language. At an early stage, he or she may hear (the sound equivalent of) 'i t s b e d t i m e'. At this early stage, there may be nothing that the child recognises in this sequence. Hence, we may suppose, he or she simply stores it as it is without coding or analysis. At some later point, the child hears (the sound equivalent of) 'i t s p l a y t i m e'. A search for good alignments should yield one like this:

```
0 i t s b e d   t i m e 0
  | | |         | | | |
1 i t s p l a y t i m e 1
```

Unification of matching patterns in this alignment should yield something like '%1 i t s #1' and '%2 t i m e #2' (each with a frequency of 2) and storage of the unmatched patterns should yield '%3 0 b e d #3' and '%3 1 p l a y #3' (each with a frequency of 1). The entire alignment may be encoded as '%4 %1 #1 %3 #3 %2 #2 #4'.

This is already like a simple grammar:

- With the debatable exception of 'i t s', all the words are picked out as discrete entities.
- The two patterns '%3 0 b e d #3' and '%3 1 p l a y #3' form a disjunctive class, similar to the distributional class *adjective* in the grammar of English. That they belong to the same class is signalled by their sharing of the pair of symbols '%3 ... #3'.
- The pattern '%4 %1 #1 %3 #3 %2 #2 #4' is like an abstract 'sentence' pattern that specifies '%1 i t s #1' at the beginning and '%2 t i m e #2' at the end, with a choice of '%3 0 b e d #3' or '%3 1 p l a y #3' in the middle.

There is, of course, a lot more to the learning of grammars than this, but this simple example should give the flavour of how the SP61 model may be developed for unsupervised inductive learning. The SNPR model, developed in the earlier research on language learning [21, 20], demonstrates the power of this approach.

## 9 Conclusion

This has been a necessarily brief summary of research developing the ICMAUS framework. I hope the ideas and examples that have been described will be enough to show the potential of these ideas and will encourage readers to investigate further.

## References


[1] Cover, T. M. and Thomas, J. A. (1991) *Elements of Information Theory*, New York: John Wiley.

[2] Gammerman, A. J. (Ed.) (1996) *Computational Learning and Probabilistic Reasoning*, Chichester: John Wiley.

[3] Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Francisco, California: Morgan Kaufmann.

[4] Rissanen, J. (1978) Modelling by the shortest data description *Automatica-J, IFAC* 14, 465–471.

[5] Sankoff, D. and Kruskall, J. B. (1983) *Time Warps, String Edits, and Macromolecules: the Theory and Practice of Sequence Comparisons*, Reading, MA: Addison-Wesley.

[6] Solomonoff, R. J. (1964) A formal theory of inductive inference. Parts I and II. *Information and Control* 7, 1–22 and 224–254.

[7] Wallace, C. S. and Boulton, D. M. (1968) An information measure for classification *Computer Journal* 11(2), 185–195.

[8] Wolff, J. G. (2000) Syntax, Parsing and Production of Natural Language in a Framework of Information Compression by Multiple Alignment, Unification and Search. *Journal of Universal Computer Science* 6(8), 781–829. A copy may be obtained from www.jucs.org/wolff/jucs_6_8. The three articles on which this article is based may be obtained from www.iicm.edu/wolff/1998d1, d2, d3, or from www.sees.bangor.ac.uk/~gerry/sp_summary.html#parsing

[9] Wolff, J. G. (1999a) Probabilistic reasoning as information compression by multiple alignment, unification and search: an introduction and overview. *Journal of Universal Computer Science* 5(7), 418–462. A copy may be obtained from www.jucs.org/wolff/jucs_5_7. The three articles on which this article is based may be obtained from www.iicm.edu/ wolff/1998a, b, c, or from www.sees.bangor.ac.uk/~gerry/sp_summary.html#PrbRs.

[10] Wolff, J. G. (1999b) 'Computing' as information compression by multiple alignment, unification and search. *Journal of Universal Computer Science* 5(11), 777–815. A copy may be obtained from www.jucs.org/wolff/jucs_5_11.

[11] Wolff, J. G. (1997) Causality, statistical learning and multiple alignment. Paper presented at the *UNICOM Seminar and Tutorial on Causal Models and Statistical Learning*, London, March 1997.

[12] Wolff, J. G. (1996) Learning and reasoning as information compression by multiple alignment, unification and search. In: [2, pp. 67–83]). An earlier version was presented at Applied Decision Technologies '95, Brunel University, April 1995 (Proceedings of Stream 1, *Computational Learning and Probabilistic Reasoning*, pp. 223–236).

[13] Wolff, J. G. (1995a) Computing as compression: an overview of the SP theory and system. *New Generation Computing* 13, 187–214.

[14] Wolff, J. G. (1995b) Computing as compression: SP20. *New Generation Computing* 13, 215–241.

[15] Wolff, J. G. (1994a) A scaleable technique for best-match retrieval of sequential information using metrics-guided search. *Journal of Information Science* 20(1), 16–28.

[16] Wolff, J. G. (1994b) Towards a new concept of software. *Software Engineering Journal* 9(1), 27–38.

[17] Wolff, J. G. (1993) Computing, cognition and information compression. *AI Communications* 6(2), 107–127.

[18] Wolff, J. G. (1991) *Towards a Theory of Cognition and Computing*, Ellis Horwood, Chichester.

[19] Wolff, J. G. (1990) Simplicity and power—some unifying ideas in computing. *Computer Journal* 33(6), 518–534.

[20] Wolff, J. G. (1988) Learning syntax and meanings through optimization and distributional analysis. In Y. Levy, I. M. Schlesinger and M. D. S. Braine (eds.), *Categories and Processes in Language Acquisition*, Lawrence Erlbaum, Hillsdale, NJ. Reprinted in Chapter 2 of [18].

[21] Wolff, J. G. (1982) Language acquisition, data compression and generalization. *Language & Communication* 2, 57–89. Reprinted in Chapter 3 of [18].